Biopsy in Renal Cell Carcinoma

Surveillance Epidemiology and End Results (SEER) estimates that 58,240 people will be diagnosed with cancer of the kidney and renal pelvis, while 13,040 will die in 2010\textsuperscript{1}. Renal cell cancer accounts for about two percent of all adult cancers\textsuperscript{2}. More than 80\% of cancers of the kidney arise in the renal parenchyma, while the rest arise in the pelvis\textsuperscript{3}. Historically, renal cell carcinoma (RCC) has been treated with nephrectomy, often with no tissue diagnosis, because of the adverse effects and possibility of false negatives associated with renal mass biopsy. With the advancement of imaging modalities in the last decades, smaller tumors are now being discovered that pose therapeutic dilemmas that did not exist in the past. Growing numbers of studies are showing a substantial proportion of these masses is benign on nephrectomy pathology leading urologists to question the current paradigm and reconsider biopsy\textsuperscript{4,5}. Due to improved sampling techniques and advances in immunohistochemistry, recent studies have reported higher levels of sensitivity and specificity for biopsy than were previously reported in the literature\textsuperscript{6-8}. If biopsy could be proven to be dependable, options such as active surveillance and ablative therapies could provide physicians and patients with better choices if patients are not good surgical candidates.

Etiology

Though the exact mechanisms of carcinogenesis remain uncertain, tobacco and obesity are the two biggest risk factors for renal cell carcinoma (RCC), associated with 20\% and 30\% of cases, respectively\textsuperscript{9,10}. Another important risk factor is hypertension, which has been proven to act via an independent mechanism from obesity\textsuperscript{9}. While initial studies suggested a link between analgesics and RCC, recent studies with acetaminophen, aspirin, and NSAIDs showed no
significant relationship to cancer. This may be due to the fact that initial studies considered phenacetin, which has now been off the market for over 25 years\textsuperscript{9}. Studies have failed to show a significant association between diet and RCC\textsuperscript{9}.

**Tumor Characteristics**

In 2002, the International Union Against Cancer (UICC) updated the TNM staging system for renal cell carcinoma by further classifying the T1 group into T1a and T1b; T1a refers to tumors within the kidney that are less than or equal to four centimeters, while T1b refers to tumors within the kidney that are between four and seven centimeters\textsuperscript{11}. Stage II cancer is larger than 7 centimeters but is still limited to the kidney. Stage III refers to tumors that have spread to one lymph node or into the fatty tissue or veins around the kidney, but not to any distant organs. Lastly, Stage IV cancer involves spread to more than one lymph node, a lymph node not near the kidney, or to distant sites\textsuperscript{12}. With regards to RCC grading, Fuhrman et al. presented a system based on 100 patients after nephrectomy in which four nuclear grades were defined based on size, irregularity and nucleolar prominence. Though many alternate grading schemes have been presented, the Fuhrman grading system remains the most widely used system in the United States\textsuperscript{12}.

**Changes in Epidemiology**

Incidence rates of RCC have increased steadily over the past two decades in the United States and the trend is accompanied by an increase in early stage cancer\textsuperscript{13}. Historically, renal cancer was diagnosed based on a triad of symptoms including flank pain, a flank mass, and hematuria. More recently, due to the ubiquity of modern imaging, a large proportion of renal masses are found incidentally before any symptoms are present. A 1998 paper examined cases of
RCC at Yale New Haven Hospital from 1989 and 1993 and found that more than 60% were discovered incidentally on imaging and showed this to be an increasing trend compared to previous decades. Of the 135 pathology specimens collected from the lab, 131 had complete medical records accessible. The authors referred to masses as incidental if they lacked signs or symptoms of RCC, such as flank mass, flank pain, or hematuria, or if the physician had no suspicion of RCC as noted in the medical record. Findings such as these have been attributed to increasing detection of asymptomatic tumors by imaging procedures such as ultrasound, computed tomography and magnetic resonance imaging. Unpublished data from the Health Care Financing Administration, which reported a 73% rise in abdominal/pelvic CT scans or MRIs to Medicare beneficiaries, supports this hypothesis.

The increased detection is coupled with a decrease in the average size of masses at diagnosis due to the advancing precision of imaging modalities; a recent study reported a decrease in mean tumor size from 4.1 to 3.6 cm between 1993 and 2004.

Based on the idea that prompt removal of tumors leads to better survival outcomes, an increase in the rate of nephrectomy has mirrored the increase in incidence of renal masses. A paper by Hollingsworth et al. collected information on 40,813 cases of RCC from nine SEER areas and found that from 1983 to 2002, the overall incidence of RCC increased from 7.1 to 10.8 cases per 100,000 people, an increase of 52%. The largest increase was observed in tumors between two and four centimeters. For each tumor size category, there was a similar increase in the rate of renal surgery in the same time period. These data showed a substantially higher increase in incidence than a similar paper using SEER data by Chow et al. which showed between a 2.3% and 4.3% increase in renal cell cancer incidence between 1975-1977 and 1993-1995. Despite increased rates of surgical treatment, the study found that RCC-specific mortality
rates rose from 1.2 to 3.2 deaths per 100,000 people, with an increase seen in all tumor size groups. Although five-year survival for RCC has improved in that time period, the overall mortality has increased, which the authors attribute to lead and length time bias. Because smaller tumors are being detected, the five-year survival percentages are increasing, but overall survival is not. The authors argue that despite earlier detection of masses and increased rates of surgery, population data do not show a decrease in mortality, which argues for “a reassessment of the current treatment paradigm.” As the authors admit, the study does have limitations. Among them, SEER does not collect patient comorbidity data and thus it is impossible to know the causes of mortality in all patients. Additionally, over 15% of patients were missing tumor size information and were excluded from the study. 

**Imaging**

Because of the increased accessibility and quality of imaging, renal masses are now routinely found on CT scans and MRIs. When examining renal masses on imaging, radiologists assess calcification, pattern (homogenous vs. heterogenous), attenuation, septations, wall characteristics, contrast enhancement, and enhancement patterns. While many findings, such as contrast enhancement and noncalcification, are more common in carcinomatous lesions, benign lesions have also been shown to have these features. Thus, many authors have concluded that despite substantial advancements in technology and resolution, imaging alone cannot distinguish between malignant and benign masses with certainty. A study from 2004 examined 162 solid renal masses that were initially evaluated using imaging, and found that 145 were renal cell carcinomas and 17 were benign. Of the 17 benign cases, 16 of the patients underwent some sort of surgical procedure. Despite recent efforts to increase the accuracy of imaging, traditional findings such as a homogenous hypervascular appearance or a spoke-wheel
on angiography have been shown to be neither sensitive nor specific\textsuperscript{19}. Although recent advances in imaging allow accurate differentiation between cystic and solid lesions, differentiating benign and malignant renal masses remains largely unreliable\textsuperscript{20,21}.

Imaging can be diagnostic, however, in one specific type of renal mass: angiomyolipomas, also known as renal hamartomas. According to studies by Bosniak, attenuation of fat on CT scans in angiomyolipomas can be diagnostic in more than 90\% of these masses\textsuperscript{22 23}. However, when angiomyolipomas have no fat present on imaging, they are indistinguishable from renal cell carcinoma and merit further investigation\textsuperscript{24}.

**Treatment**

Despite substantial advances in oncologic treatment, renal cancer still remains resistant to standard chemotherapeutic agents. However, low stage RCC can be successfully treated with extirpative surgery, with better prognosis for those with lower stage disease\textsuperscript{25}. Historically, radical nephrectomy, with wide excision of the kidney outside of Gerota’s fascia to include the adrenal gland and perirenal fat, was preferred because of concern that extrarenal involvement lead to surgical failure\textsuperscript{26}. However, nephron-sparing surgery, initially performed in those with a solitary kidney or pre-existing renal insufficiency with the hopes of preserving renal function, has become increasingly popular in recent decades.

In a 2000 paper, Lau et al. conducted a retrospective analysis of radical nephrectomy versus nephron-sparing surgery. The patients were matched on age, sex, tumor size, pathologic tumor stage and grade, and year in which the surgery was performed. The study showed that over a 15 year follow-up, 77\% of those undergoing radical nephrectomy were alive with no evidence of disease compared with 79\% of nephron-sparing surgery patients. At ten year follow-up, there
were similar rates of contralateral recurrence and metastasis. The cumulative incidence of chronic renal insufficiency (defined as greater than 2.0 mg/dL at least 30 days after surgery) was 22.4% and 11.6% in the radical nephrectomy group and nephron-sparing surgery group, respectively. The authors suggest that nephron-sparing surgery is as effective as radical nephrectomy from an oncologic standpoint and may potentially have better outcomes with respect to complications. Compared to earlier studies in which nephron-sparing surgery was only done in those with solitary kidneys or pre-existing renal insufficiency, the study enrolled patients with single unilateral RCC and a normal contralateral kidney, which strengthens generalizability. These findings coincided with those from a multi-center Austrian study as well as a single institution German study showing comparable cancer-free survival in nephron-sparing surgery versus radical nephrectomy\textsuperscript{27,28}.

However, due to its non-randomized and retrospective design, the conclusions from this study are limited. Despite matching on tumor size, the average size in the radical nephrectomy group was 3.7 cm compared with 3.3 cm in the nephron-sparing surgery group, with the difference being statistically significant. It is likely that the smaller size of tumors in the nephron-sparing surgery group could have led to improved outcomes both in terms of oncologic outcome and renal function. It is also impossible to match and adjust for all the potential factors that could affect survival and renal function retrospectively. A prospective randomized control trial addressed this issue in a study that recruited patients with small (\(\leq 5\) cm) solitary, T1-T2 N0 M0 masses and a normal contralateral kidney. The authors randomized patients to nephron sparing surgery or radical nephrectomy and analyzed overall survival, with disease-specific survival, progression, and surgical side effects as secondary end-points. The two groups were similar with regard to age, tumor size, tumor stage, tumor grade, and WHO performance status.
The intention-to-treat analysis showed a 10-year overall survival rate of 75.7% for NSS and 81.1% for RN, with an estimated hazard ratio of 1.50 (95% CI 1.03-2.16). However, this number became insignificant when the analysis was limited to those with actual renal cell carcinoma rather than surgery performed for benign masses. The hazard ratio of death from renal cancer was not significantly higher in the NSS group, nor was progression of cancer. A major weakness of the study was that the authors reported recruiting only 541 patients due to poor accrual despite calculating a necessary sample size of 1300 patients to detect a proposed 3% difference in 5-year survival. Additionally, the authors reported that 5.9% of the patients randomized to radical nephrectomy underwent nephron sparing surgery and 14.6% of patients randomized to nephron sparing surgery underwent radical nephrectomy. The finding that overall survival is slightly better in RN than NSS is in contrast with many previously published studies on this topic. However, those studies are limited by their retrospective design, heterogeneous stage of tumors between groups, and lack of renal function outcomes. As far as we know, this is the only randomized trial to date on this subject and more studies like this need to be done in order to definitively answer this question.

In 2007, Novick and Derweesh outlined absolute, relative, and elective indications for nephron-sparing surgery compared to radical nephrectomy. Absolute indications apply to situations in which nephrectomy would result in an immediate need for dialysis. Such situations include patients with a solitary kidney, bilateral renal cell carcinoma, or patients in whom the contralateral kidney cannot sustain normal kidney function. Relative contraindications apply to situations in which the long-term viability of the remaining kidney is in doubt, such as stone disease, chronic pyelonephritis, renal artery stenosis, or systemic diseases such as diabetes.
Elective indications are for those patients with unilateral renal cell carcinoma less than 4 cm and a normal contralateral kidney\textsuperscript{26}.

**Benign Findings on Nephrectomy**

For tumors smaller than 4 cm, the gold standard for treatment is radical or nephron sparing surgery\textsuperscript{30}. However, a paper from 1987 showed that roughly 15\% of renal masses detected on CT scan were benign\textsuperscript{31}. More recent papers have shown even higher rates of benign findings on nephrectomy pathology. In a paper published in 2003, Lane and Gill retrospectively examined patients who underwent laparoscopic partial nephrectomies at the Cleveland Clinic and found 56 patients with at least five years of follow-up. Though the purpose of the study was to report oncologic outcomes after partial nephrectomy, the data showed a benign diagnosis on final pathology in 19 of the 56 patients (38\%)\textsuperscript{32}. These numbers may not provide an accurate representation of the proportion of benign masses because some of the masses were symptomatic; in a population of asymptomatic incidental renal masses, it is possible that the proportion of benign masses would be even higher. A similar paper in 2006 found that 38 of 123 (31\%) masses removed by laparoscopic partial nephrectomy were found to be benign on final pathology. The average size of the masses in the study was 2.6 cm, which is promising for generalizability to small renal masses, which are generally considered to be less than 4 cm. The indications for nephrectomy, however, were difficult to ascertain as the authors only reported performing surgery for “enhancing renal masses\textsuperscript{33}.” In 2007, Gill et al. reported their findings of 1800 laparoscopic and open partial nephrectomies, showing that 21.4\% of nephrectomy specimens showed benign histology\textsuperscript{5}. 
The proportion of nephrectomies for benign masses is even more disconcerting when stratified by size, as shown in a case series by Frank et al. that examined 2770 nephrectomies due to solid renal masses and found that those less than 3 cm were 25% benign, those less than 2 cm were 30% benign and those less than 1 cm were 44% benign\textsuperscript{34}. With the average age of RCC being 65 years old, many patients undergoing procedures are at higher risk for post-operative complications or mortality due to various comorbidities\textsuperscript{30}. Because of this increased risk, active surveillance and modern ablative therapies are likely a better option in this population. Biopsy, if proven to be accurate and dependable, could assist the physician and the patient in making a more informed decision. Even young patients may have chronic renal insufficiency, solitary kidneys, or other medical conditions that seriously complicate surgery; thus, watchful waiting would be a welcome option to many patients if it could be proven as a safe alternative\textsuperscript{35}.

**The resurgence of biopsy**

Benign findings on nephrectomy and the known indolence of small malignancies coupled with the emergence of newer treatments such as radio-frequency ablation and cryoablation, which require pre-procedure biopsy, have caused physicians to reconsider the role of renal mass biopsy\textsuperscript{19}. Supporters of biopsy claim that it is safe, that the majority of biopsied lesions are benign, and that biopsy can decrease unnecessary surgery\textsuperscript{34}. Opponents argue that biopsy carries too high a risk of false negatives\textsuperscript{36}, can lead to biopsy track seeding with cancer cells, and leads to complications\textsuperscript{37}.

While bleeding when the renal capsule is punctured is to be expected with biopsy, one case series showed that renal hemorrhage resulting in hospitalization or blood transfusion occurred in 1% to 2% of biopsies\textsuperscript{38}. A more recent series in which a CT was taken immediately
after biopsy to check for complications, the authors reported no patients in whom a hematoma was found or any patients requiring intervention related to biopsy complications. In a meta-analysis by Lane et al., the authors observed complications in renal mass biopsies in seven case series done after 2001. Out of a total of 362 biopsies, one series in the review showed one major complication from biopsy while the remaining six showed zero major complications. There were 17 minor complications reported in the review. An important consideration is that these are case series reported from academic centers, and that their characteristics may not be applicable to all types of hospitals or urology practices. The authors did not report skills levels of surgeons or the volume of cases that these centers are exposed to. Additionally, the papers did not adjust for baseline status of the patients and thus conclusions are limited.

A case report of biopsy needle track seeding with malignant cells from a study in 1991 raised alarms about the risk of percutaneous biopsy. Recently, however, reports of tumor seeding have been exceedingly rare, in part because of the use of the coaxial system. This approach uses a larger gauge introducer which then remains in place while a smaller needle can make multiple passes to obtain tissue. This approach has been reported to reduce the chance of tumor seeding because it limits the number of times the renal capsule is penetrated. Numerous reports with follow-up periods ranging from one to five years have shown no evidence of needle track seeding. The main argument against these findings is that the follow-up period may not be long enough to detect possible tumor seeding.

Not only have the rates of complications and tumor seeding been shown to be lower than suspected, but the historically poor numbers for biopsy sensitivity and specificity may be more complicated than previously considered as well. Historically, studies have regarded inconclusive or failed biopsies as false negatives when calculating sensitivity and specificity for renal mass
biopsy. For example, in an American Urology Association Update Series from 2000, the authors report sensitivity for core biopsy of renal masses as low as 70% and state that “false negatives are most often due to insufficient specimens which are often bloody aspirates.” Based on these reports, they conclude that biopsy is not indicated for a solitary renal lesion that looks like RCC on imaging. However, taking these biopsies out of the denominator can portray a drastically different situation depending on the frequency of biopsy failure. In more recent studies, where inconclusive biopsies, i.e. those showing normal renal parenchyma, necrotic tissue, or hemorrhage are excluded from the analysis, the accuracy of biopsy in deciphering benign versus malignant masses has been reported between 89% and 98%.

Because false negatives based on an adequate sample have been shown to be less common, it is possible that the rising sensitivity values seen in recent studies are attributed not to improvements in techniques, but to improvements in denominators.

Regardless of the classification of non-diagnostic biopsies, multiple studies have shown that the use of biopsy decreases the number of surgeries performed for renal masses, with Wood et al. reporting a 41% decrease in the number of surgeries and Somani et al. reporting 24% of patients avoiding surgery. In light of new information and changing opinions, this paper aims to examine the literature for the sensitivity and specificity of renal mass biopsy in those with incidentally detected small renal masses and to report the frequency of adverse events including bleeding episodes and needle track seeding.
Methods

This systematic review included published studies describing different measures of performance for renal mass biopsy in imaging-detected small renal masses. Biopsy of renal masses is currently not routine practice and thus the studies were limited to academic medical centers participating in research. Because of recent advances in immunohistochemistry and pathology techniques, we sought to perform a review of recent articles and chose 2005 as a cutoff. In March 2011, we identified prospective or retrospective cohort studies and case series by searching the MEDLINE and EMBASE databases from January 1, 2005 to March 1, 2011. We also hand searched reference lists. A research librarian at the University of North Carolina Health Sciences Library assisted in creation of the search strategy.

We searched the following terms in MEDLINE:


[majr] = Major MeSH term

[mh] = MeSH term

[sh] = Subheading

We searched the following terms in EMBASE:
“kidney tumor'/exp/mj AND biopsy'/exp/mj AND ('predictive value of tests' OR 'sensitivity and specificity' OR 'reproducibility of results' OR 'retrospective studies' OR 'risk assessment' OR 'prognosis' OR 'incidental findings' OR 'diagnosis' OR 'classification' OR 'pathology')

Limits: Humans; English”

Two readers independently first reviewed all titles identified by both searches excluding those about non-renal malignancies, editorial or “letter” articles, studies not discussing biopsy, and individual patient case reports. The readers then read the remaining abstracts excluding those in which renal masses over 5 cm were reported, articles reported ex-vivo biopsies, articles reported laparoscopic biopsies, and articles that reported biopsy before radiofrequency ablation. All remaining articles were fully reviewed.

In the full review, we sought prospective or retrospective cohort studies and case series in which patients underwent renal mass biopsy for a renal mass detected on imaging and suspected to be renal carcinoma. Because the average age at diagnosis of renal cell carcinoma is 65 years old, we considered patients over age 40 in order to include as many patients as possible. It was important to have an age minimum in order to prevent pediatric renal cancers from being included in the review, because their pathology and epidemiology are quite different from adult renal cancers. We did not discriminate based on biopsy technique, i.e. FNA versus core biopsy, or on needle size. We did not discriminate based on nephrectomy technique, i.e. open radical nephrectomy, laparoscopic radical nephrectomy, open partial nephrectomy, or laparoscopic partial nephrectomy. Our primary outcomes were sensitivity and specificity of biopsy findings compared to ex vivo pathology findings after nephrectomy, rate of adverse events, and findings of needle track seeding. Secondary outcomes were sensitivity and specificity for RCC subtype and Fuhrman grade, and percentage of diagnostic or “conclusive” biopsies.
Population | Adults over age 40 with incidentally detected renal mass found on imaging
---|---
Intervention | Image-guided biopsy
Comparison | If nephrectomy specimen available, compare biopsy findings to nephrectomy pathology
| If no nephrectomy, consider 2 year symptom-free surveillance of a benign biopsy correct
Outcome | Benign versus malignant biopsy result
Time | At least a two-year follow-up of cases under surveillance
Setting | Academic medical centers
Study Type | Randomized control trials, cohort studies, case series

Table 1. USPSTF criteria for cohort studies.

**Quality Assessment Tool**

The United States Preventive Services Taskforce provides the following criteria for randomized control trials and cohort studies:

- Initial assembly of comparable groups:
  - For RCTs: adequate randomization, including first concealment and whether potential confounders were distributed equally among groups.
  - For cohort studies: consideration of potential confounders with either restriction or measurement for adjustment in the analysis; consideration of inception cohorts.
- Maintenance of comparable groups (includes attrition, cross-overs, adherence, contamination).
- Important differential loss to follow-up or overall high loss to follow-up.
- Measurements: equal, reliable, and valid (includes masking of outcome assessment).
- Clear definition of interventions.
- All important outcomes considered.
- Analysis: adjustment for potential confounders for cohort studies, or intention to treat analysis for RCTs.
Based on these criteria, the USPSTF offers the following three categories:

**Good:** Meets all criteria: Comparable groups are assembled initially and maintained throughout the study (follow-up at least 80 percent); reliable and valid measurement instruments are used and applied equally to the groups; interventions are spelled out clearly; all important outcomes are considered; and appropriate attention to confounders in analysis. In addition, for RCTs, intention to treat analysis is used.

**Fair:** Studies will be graded "fair" if any or all of the following problems occur, without the fatal flaws noted in the "poor" category below: Generally comparable groups are assembled initially but some question remains whether some (although not major) differences occurred with follow-up; measurement instruments are acceptable (although not the best) and generally applied equally; some but not all important outcomes are considered; and some but not all potential confounders are accounted for. Intention to treat analysis is done for RCTs.

**Poor:** Studies will be graded "poor" if any of the following fatal flaws exists: Groups assembled initially are not close to being comparable or maintained throughout the study; unreliable or invalid measurement instruments are used or not applied at all equally among groups (including not masking outcome assessment); and key confounders are given little or no attention. For RCTs, intention to treat analysis is lacking.
Figure 1. Search strategy and outcomes.
Results

Study Selection
The PUBMED search resulted in 369 articles, which we narrowed down to 123 based on the titles of the articles. We subsequently reviewed 123 abstracts and from these, chose 9 papers to include in the study. The EMBASE search yielded 203 articles, which we narrowed down to 39 articles based on titles. We subsequently reviewed 39 abstracts, and chose 3 papers to include in the study. There was substantial overlap between the two databases despite efforts to prevent duplication. Because we searched EMBASE after PUBMED many of the papers from the EMBASE search were already included because they arose in the PUBMED database.

FNA versus core
Eight of the studies used needle core biopsy rather than fine needle aspiration, however three studies did use both methods. In the study by Veltri et al., interventional radiologists only used FNA technique when a pathologist was present in the room.

CT/US guided
While many studies included both computed tomography and ultrasound-guided biopsy, very few presented details on the performance of one compared with the other. Nine of the studies reported use of both CT and ultrasound, while one study reported ultrasound-guided only, and another reported CT only. Only Volpe et al. reported an analysis of one method versus the other and found that there was no statistical significance between the accuracy of ultrasound-guided biopsy versus CT-guided biopsy. There was also no statistically significant improvement when using both methods.30

Adequate tissue for diagnosis
One of the central arguments against biopsy has been that it often yields inadequate tissue for diagnosis. While we did not search for this specifically, all of the studies included in the analysis provided this information. The values ranged from as low as 68.5% adequate biopsy in the Vasudevan et al. study to 100% adequate biopsy in the Lebret et al. study. All of the studies used 18 gauge needles to perform core biopsy, with the exception of the Somani et al. study using 16 and 18 gauge needles and the Vasudevan study using 16 gauge needles only. No stratification was provided based on needle size in the Somani study, but it is interesting that the lowest rate of successful biopsies was reported in the study using exclusively the largest gauge needle\textsuperscript{48}. In the studies using FNA technique, one did not report needle size\textsuperscript{49}, one used 22 gauge needles\textsuperscript{30} and another reported 21-22 gauge use\textsuperscript{50}.

Many studies described normal renal parenchyma (i.e. not the mass), necrotic or hemorrhagic areas, inflammation, fibrosis, or specimens in which the tissue origin could not be determined. Depending on the study, the numbers are derived from either the first biopsy taken, with subsequent re-biopsy success rates reported, or in many studies, patients went on to repeat biopsy if the first sample was inadequate and these numbers are included in the total percentage of adequate biopsies. The Schmidbauer et al. study reported two different rates of adequate biopsy, 89% FNA, and 97% for core biopsy. For FNA, failed biopsies reported no cells in one sample and blood cells only in the remaining samples, while core biopsy showed normal renal parenchyma. This study eventually stopped using FNA after the first 44 patients.

**Sensitivity for Benign versus Malignant**

Despite our initial interest in sensitivity and specificity, only two studies reported these values and the other nine reported “accuracy” values. In these studies, accuracy for distinguishing
benign versus malignant pathology ranges from 92% to 100% in those with adequate biopsy. The Veltri et al. study reported accuracy for FNA alone as 88.4%, core biopsy alone as 94.8%, and in those undergoing both FNA and core biopsy, they reported an accuracy of 94.8%. Maturen et al. reported a sensitivity of 85/87 (97.7%) and a specificity of 100% for diagnosing benign versus malignant pathology. The two incorrect diagnoses in the sensitivity calculation were not “benign” but rather “nondiagnostic” biopsies that turned out to be malignant on nephrectomy. In all of the studies, substantially fewer patients underwent nephrectomy than had a diagnostic biopsy. There were very few nephrectomy specimens for those with a benign biopsy. In all but one paper, accuracy was only reported for the subset of patients with a diagnostic biopsy that went on to nephrectomy, because the nephrectomy specimen could serve as a gold standard. However, Maturen et al. created criteria under which they considered an observed patient with a benign biopsy correctly diagnosed after two years of no mass growth or symptoms.

**RCC subtype**

Although not as many studies reported accuracy for RCC subtype as did for benign versus malignant, for those studies that did, the numbers ranged from 82% to 100% accurate. Schmidbauer et al. reported values stratified by FNA versus core biopsy and showed that accuracy for subtype was 89% in FNA while it was 97% in core biopsy. Two studies, one by Jaff et al. and the other by Volpe et al. both reported 100% in accuracy in RCC subtyping with sample sizes of 20 and 27, respectively. Fuhrman grading, only reported in four studies, ranged from 28% to 76% accurate depending on whether the study used FNA or core biopsy. Lebret et al. found 76% accuracy when reporting Fuhrman grade as high versus low, which may have significance with regards to further treatment.
Adverse Events

Though definitions of adverse events were not standardized, the majority of studies dichotomized adverse events into significant versus insignificant, or major versus minor groups. In most cases, major adverse events required intervention, an emergency room visit, or hospital admission. Minor events were those that could be monitored without any action. Studies differed in their post-procedure protocols, with some centers performing ultrasound on all patients who underwent biopsy, while others only took action if complications arose. No studies reported needle track seeding, a major source of concern in early studies. The Wang et al. study reported the highest rate of complications requiring intervention, at 7.3%. Six studies reported either no complications or only minor events requiring no intervention. The Vasudevan study reported one complication in which a patient needed two units of red blood cells, while Maturen et al. reported one patient who needed four units. The total number of biopsies in all the studies considered is 1242 and the total number of serious complications (emergency room visit, admission, or medical intervention) reported was 14, resulting in a rate of 1.1%.

Management of Benign Biopsy

The management of those patients with a benign biopsy varied across studies. Six of the studies reported no scenarios in which patients with a benign biopsy underwent nephrectomy and even in the five studies that did, the numbers were all in the single digits. Average follow-up of nonsurgical patients ranged from 9.7 months in the Maturen et al. study to 37 months in the Veltri et al. study. The most comprehensive definition of benign biopsy came from the Maturen et al. study, in which the authors defined “definitely benign” as those with no growth at two years and “probably benign” as those with no growth at six months to two years. The majority of
studies reported surveillance regimens using CT or MRI every six months, with some spacing out imaging to once a year after a certain amount of time with no growth. The Shannon et al. study reported growth of monitored lesions, but reported no symptoms or subsequent surgery for the masses.

**Change in Management**

Because eight of the eleven studies were retrospective, the authors reported results as if they would have changed management. These changes referred to patients avoiding surgery based on biopsy diagnosis of benign lesions, diagnosis of low grade lesions, or diagnosis of metastases that required treatment of the initial cancer rather than nephrectomy. In other cases, the biopsy results affected the type of surgery (i.e. nephron sparing surgery versus radical nephrectomy) or led physicians to use radio frequency ablation techniques rather than extirpative surgery. Even in cases where surveillance was planned initially, a benign biopsy led to less strict surveillance protocols. Eight of the studies reported the percentage of patients with a change in a management, with numbers ranging from 15% to 69% of patients affected.
<table>
<thead>
<tr>
<th>Study</th>
<th>Initial Assembly of Comparable Groups</th>
<th>Differential or high loss to follow-up</th>
<th>Measurement</th>
<th>Clear definition of intervention s</th>
<th>Important outcomes considered</th>
<th>Adjustment for potential confounders</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang 32</td>
<td>no</td>
<td>3% benign biopsies lost to follow-up. 0% of malignant</td>
<td>Details of IHC-unknown. Biopsy pathology compared to nephrectomy pathology</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>poor</td>
</tr>
<tr>
<td>Volpe 30</td>
<td>no</td>
<td>No reported loss to follow-up</td>
<td>Details of IHC-yes. Biopsy pathology compared to nephrectomy pathology</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>poor</td>
</tr>
<tr>
<td>Shannon 31</td>
<td>no</td>
<td>7/28 patients with no diagnosis lost to follow-up. 0 patients with diagnostic biopsy lost to follow-up</td>
<td>Details of IHC-unknown. Biopsy pathology compared to nephrectomy pathology</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>poor</td>
</tr>
<tr>
<td>Schmidbauer 29</td>
<td>no</td>
<td>No report of loss to follow-up</td>
<td>Details of IHC-unknown. Biopsy pathology compared to nephrectomy pathology</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>poor</td>
</tr>
<tr>
<td>Lebret 31</td>
<td>no</td>
<td>No report of loss to follow-up</td>
<td>Details of IHC-yes. Biopsy pathology compared to nephrectomy pathology</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>poor</td>
</tr>
<tr>
<td>Reichelt 34</td>
<td>no</td>
<td>No report of loss to follow-up</td>
<td>Details of IHC-yes. Biopsy pathology compared to nephrectomy pathology</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>poor</td>
</tr>
<tr>
<td>Maturen 26</td>
<td>no</td>
<td>No report of loss to follow-up</td>
<td>Details of IHC-unknown. Biopsy pathology compared to nephrectomy pathology</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>poor</td>
</tr>
<tr>
<td>Somani 47</td>
<td>no</td>
<td>1/27 in biopsy</td>
<td>Details of IHC-unknown. Biopsy pathology compared to nephrectomy pathology</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>poor</td>
</tr>
<tr>
<td>Study</td>
<td>Biopsy</td>
<td>Loss to follow-up</td>
<td>IHC Details</td>
<td>Quality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>-------------------</td>
<td>-------------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vasudevan</td>
<td>no</td>
<td>No report of loss to follow-up</td>
<td>Details of IHC-unknown. Biopsy pathology compared to nephrectomy pathology</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>poor</td>
</tr>
<tr>
<td>Veltri</td>
<td>no</td>
<td>12/150 biopsies lost to follow-up, no further details provided</td>
<td>Details of IHC-unknown. Biopsy pathology compared to nephrectomy pathology</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>poor</td>
</tr>
<tr>
<td>Jaff</td>
<td>no</td>
<td>No report of loss to follow-up</td>
<td>Details of IHC-unknown. Biopsy pathology compared to nephrectomy pathology</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>poor</td>
</tr>
</tbody>
</table>

Table 2. Quality criteria adapted from USPSTF cohort study quality criteria.  
*important outcomes: findings on biopsy, findings on nephrectomy (if available), adverse events, tumor seeding, death  
**IHC- immunohistochemistry
<table>
<thead>
<tr>
<th>Study</th>
<th># biopsies</th>
<th>Mean tumor size</th>
<th>Mean Age (years)</th>
<th>% Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang</td>
<td>110</td>
<td>2.7 cm</td>
<td>60.4</td>
<td>68%</td>
</tr>
<tr>
<td>Volpe</td>
<td>100</td>
<td>2.4 cm</td>
<td>60 (median)</td>
<td>unknown</td>
</tr>
<tr>
<td>Shannon</td>
<td>235</td>
<td>2.9 cm (median)</td>
<td>64</td>
<td>unknown</td>
</tr>
<tr>
<td>Schmidbauer</td>
<td>122</td>
<td>4 cm</td>
<td>63</td>
<td>80%</td>
</tr>
<tr>
<td>Lebret</td>
<td>119</td>
<td>3.3 cm</td>
<td>60 (median)</td>
<td>unknown</td>
</tr>
<tr>
<td>Reichelt</td>
<td>30</td>
<td>2.9 cm</td>
<td>63</td>
<td>60%</td>
</tr>
<tr>
<td>Maturen</td>
<td>152</td>
<td>4.1 cm</td>
<td>60</td>
<td>46%</td>
</tr>
<tr>
<td>Somani</td>
<td>70</td>
<td>unknown</td>
<td>63.8</td>
<td>61%</td>
</tr>
<tr>
<td>Vasudevan</td>
<td>100</td>
<td>unknown</td>
<td>62</td>
<td>unknown</td>
</tr>
<tr>
<td>Veltri</td>
<td>150</td>
<td>3.4 cm</td>
<td>64.5</td>
<td>63%</td>
</tr>
<tr>
<td>Jaff</td>
<td>54</td>
<td>3.3 cm</td>
<td>72 (median)</td>
<td>60%</td>
</tr>
</tbody>
</table>

Table 3. Study Characteristics

<table>
<thead>
<tr>
<th>Study</th>
<th>FNA vs core</th>
<th>Gauge Needle</th>
<th>CT/US guided</th>
<th>% adequate tissue for dx</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang</td>
<td>core</td>
<td>18 gauge</td>
<td>66 CT, 44 US</td>
<td>100/110 (90.9%)</td>
</tr>
<tr>
<td>Volpe</td>
<td>both</td>
<td>18 core</td>
<td>45 US, 11 CT,</td>
<td>84/100 (84%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>22 FNA</td>
<td>44 both</td>
<td></td>
</tr>
<tr>
<td>Shannon</td>
<td>core</td>
<td>18 gauge</td>
<td>both</td>
<td>184/235 (78.3%)</td>
</tr>
<tr>
<td>Schmidbauer</td>
<td>both</td>
<td>18 core</td>
<td>CT</td>
<td>89% in FNA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FNA not specified</td>
<td>97% in core</td>
<td></td>
</tr>
<tr>
<td>Lebret</td>
<td>core</td>
<td>18 gauge</td>
<td>both</td>
<td>79%</td>
</tr>
<tr>
<td>Reichelt</td>
<td>core</td>
<td>18 gauge</td>
<td>US</td>
<td>25/30 (83%)</td>
</tr>
<tr>
<td>Maturen</td>
<td>core</td>
<td>18 gauge</td>
<td>76 US, 76 CT</td>
<td>146/152 (96%)</td>
</tr>
<tr>
<td>Somani</td>
<td>core</td>
<td>16-18 gauge</td>
<td>both</td>
<td>61/70 (87.1%)</td>
</tr>
<tr>
<td>Vasudevan</td>
<td>core</td>
<td>16 gauge</td>
<td>both</td>
<td>63/92 (68.5%)</td>
</tr>
<tr>
<td>Veltri</td>
<td>both</td>
<td>Core not specified</td>
<td>145 US, 5 CT</td>
<td>Not reported</td>
</tr>
<tr>
<td>Jaff</td>
<td>core</td>
<td>18 gauge</td>
<td>6 US, 48 CT</td>
<td>46/54 (85.1%)</td>
</tr>
</tbody>
</table>

Table 4. Study Characteristics Continued
<table>
<thead>
<tr>
<th>Study</th>
<th>Accuracy Benign vs Malig</th>
<th>Accuracy RCC subtype</th>
<th>Accuracy for Fuhrman grade</th>
<th>Adverse Events</th>
<th>Change in Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang</td>
<td>34/34 (100%)</td>
<td>28/29 (96.6%)</td>
<td>-</td>
<td>8/110= 7%</td>
<td>Not reported</td>
</tr>
<tr>
<td>Volpe</td>
<td>20/20 (100%)</td>
<td>20/20 (100%)</td>
<td>8/12 (66.7%)</td>
<td>none</td>
<td>43/100 (43%)</td>
</tr>
<tr>
<td>Shannon</td>
<td>108/108 (100%)</td>
<td>106/108 (98%)</td>
<td>-</td>
<td>2/108 (0.9%)</td>
<td>62/235 (26%)</td>
</tr>
<tr>
<td>Schmidbauer</td>
<td>95.2% (core sens)</td>
<td>91% core</td>
<td>28% FNA 76% core</td>
<td>None requiring intervention</td>
<td>19/122 (15.6%)</td>
</tr>
<tr>
<td></td>
<td>100% (core spec)</td>
<td>86% FNA</td>
<td></td>
<td>none</td>
<td></td>
</tr>
<tr>
<td></td>
<td>90.6% (FNA sens)</td>
<td></td>
<td></td>
<td>62/235 (26%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100% (FNA spec)</td>
<td></td>
<td></td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>Lebret</td>
<td>Not presented</td>
<td>86%</td>
<td>46% 76% if high vs low</td>
<td>none</td>
<td>31/102 (30%)</td>
</tr>
<tr>
<td>Reichelt</td>
<td>17/17 (100%)</td>
<td>-</td>
<td>-</td>
<td>1 renal hematoma</td>
<td>Not reported</td>
</tr>
<tr>
<td>Maturen</td>
<td>97.7% (sensitivity)</td>
<td>-</td>
<td>-</td>
<td>2 post- 90/152 (60.5%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100% (specificity)</td>
<td></td>
<td>procedure hematomas, 1 needing 4 units pRBC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Somani</td>
<td>32/32 (100%)</td>
<td>-</td>
<td>-</td>
<td>1 requiring admission</td>
<td>17/70 (24%)</td>
</tr>
<tr>
<td>Vasudevan</td>
<td>38/38 (100%)</td>
<td>-</td>
<td>-</td>
<td>1 pt needed 2 units blood</td>
<td>Not reported</td>
</tr>
<tr>
<td>Veltri</td>
<td>119/129 (92.2%)</td>
<td>-</td>
<td>-</td>
<td>8/150 (5.3%)</td>
<td>89/150 (68.9%)</td>
</tr>
<tr>
<td>Jaff</td>
<td>100%</td>
<td>100%</td>
<td>75%</td>
<td>none</td>
<td>32/46 (69.6%)</td>
</tr>
</tbody>
</table>

Table 5. Review findings for benign vs. malignant, RCC subtype, Fuhrman grade, adverse events and change in management.

*no cases of needle track seeding were reported
<table>
<thead>
<tr>
<th>Study</th>
<th>Study Design</th>
<th>Inclusion Criteria</th>
<th>Blinding of pathologists?</th>
<th>Details of immunohistochemistry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang</td>
<td>retrospective case series</td>
<td>SRM ≤ 4 cm. Adults over 18, excluded cystic masses</td>
<td>unknown</td>
<td>unknown</td>
</tr>
<tr>
<td>Volpe</td>
<td>retrospective case series</td>
<td>Incidental, isolated SRM ≤4 cm.</td>
<td>unknown</td>
<td>yes</td>
</tr>
<tr>
<td>Shannon</td>
<td>retrospective case series</td>
<td>Incidental, asymptomatic mass under 5 cm</td>
<td>unknown</td>
<td>unknown</td>
</tr>
<tr>
<td>Schmidbauer</td>
<td>prospective case series</td>
<td>Solid renal masses, cystic excluded</td>
<td>unknown</td>
<td>unknown</td>
</tr>
<tr>
<td>Lebret</td>
<td>retrospective case series</td>
<td>SRM ≤ 4 cm or those missing radiologic criteria</td>
<td>unknown</td>
<td>yes</td>
</tr>
<tr>
<td>Reichelt</td>
<td>prospective case series</td>
<td>Noncystic, homogenous mass found on US</td>
<td>unknown</td>
<td>yes</td>
</tr>
<tr>
<td>Maturen</td>
<td>retrospective case series</td>
<td>-</td>
<td>unknown</td>
<td>unknown</td>
</tr>
<tr>
<td>Somani</td>
<td>prospective case series</td>
<td>Biopsied those in which &quot;it was not possible to characterize renal masses as either malignant or benign on imaging characteristics alone&quot;</td>
<td>unknown</td>
<td>yes</td>
</tr>
<tr>
<td>Vasudevan</td>
<td>retrospective case series</td>
<td>Asymptomatic, incidental masses under 5 cm</td>
<td>unknown</td>
<td>unknown</td>
</tr>
<tr>
<td>Veltri</td>
<td>retrospective case series</td>
<td>Radiology did not provided sufficient diagnosis</td>
<td>unknown</td>
<td>unknown</td>
</tr>
<tr>
<td>Jaff</td>
<td>retrospective case series</td>
<td>Solitary kidney, RFA candidate, suspect oncoytoma, lymphoma or mets, high risk for surgery</td>
<td>unknown</td>
<td>unknown</td>
</tr>
</tbody>
</table>

Table 6. Study Description

*SRM= solitary renal masses
Discussion

The papers we reviewed claim promising results for the use of biopsy in lesions smaller than four centimeters. In these papers, biopsy was effective at distinguishing benign versus malignant masses; in the papers that reported it, biopsy could predict RCC subtype with accuracy in the 80-100% range. None of the papers reported needle track seeding, and major complications were rare. These findings coincide with other recent literature reviews on the subject that found improved sensitivity and specificity of renal mass biopsy and reported low rates of complications and no needle track seeding\(^1\text{9, 56}\). Two recent studies found that Fuhrman grade was correctly interpreted in 70\(^4\text{4}\) and 83\(^7\) of cases, which also coincides with our findings.

However, an important distinction must be made. As proposed in reviews by Lane et. al in 2008 and Samplaski et al. in 2010, the idea that biopsy failures should not be counted as false negatives is a departure from past literature\(^4\text{6, 57-62}\). Changing these denominators can give substantially higher values for sensitivity depending on the rate of biopsy failure, which may not be an entirely accurate representation of the data. By not counting these patients, some argue that the numbers ignore the non-trivial process of going through biopsy after which there is no additional information. While this is true, the rate of complications in our study is very low. Additionally, we argue that clinicians and pathologists know that a failed biopsy does not indicate benign status and would not make patient care decisions under this assumption. Thus, an important category for studies on this topic to include is “percentage/proportion of adequate biopsy.” Using this model, sensitivity and specificity calculations can be supplemented by information about the frequency of biopsy failure. We do worry however, that the increase in sensitivity reported in our review and other recent reviews compared to previous studies may be
due in part to this method of reporting; a true comparison to previous studies is difficult because of this change in the denominator.

Even if renal mass biopsy is safer and more accurate than previously thought, questions have arisen about the costs incurred by this approach. A cost-effectiveness analysis by Pandharipande, et al. showed that the risks that have kept physicians from using renal biopsy are “at least equaled by those risks incurred by performing empiric surgery in all patients.” The Markov model developed in the paper used a base-case of a 65 year old man with an incidentally detected tumor under four centimeters who could undergo empiric surgery or tumor surveillance. The base-case sensitivity and specificity were 0.9 and 1, respectively, but sensitivity analyses for both values ranged from 0.5 to 1. By using a quasi-societal model, the authors did not factor in time costs to the patient. Using the base-case model, the biopsy strategy yielded a four day longer life expectancy with a $3466 lower lifetime cost than empiric nephron sparing surgery. The biopsy strategy also dominated across the majority of sensitivity ranges during a one-way sensitivity analysis. Even when the mortality of nephron sparing surgery was assumed to be 0%, the empiric surgery model did not dominate the biopsy model. However, when biopsy sensitivity fell to less than 0.78, the surgical model dominated the biopsy model. Based on findings from the studies that we reviewed, the biopsy model should dominate in the current environment.

Limitations of the study should not nullify its findings, but they are worth considering. For example, because there is limited data on surveillance studies, the authors concede that it was difficult to model biopsy negative tumor outcomes. The authors also note that there are no accepted guidelines for surveillance, which creates difficulty in making assumptions for the model.

Limitations of Current Research
Despite increased interest in the area of renal mass biopsy, the current research still has many limitations. Based on USPSTF quality criteria, all of the papers in our study earned a *poor* quality rating. The majority of the studies are retrospective case series from major academic medical centers, in which patient outcomes are reported from medical records that are often lacking demographic information and long-term follow-up. All of the studies lacked randomization and control groups, and no study adjusted for baseline status of patients or other possible confounding variables. Additionally, many studies do not elaborate on inclusion criteria for biopsy. For the purposes of our research question, it is important that patients undergo biopsy for an asymptomatic incidental renal mass, because this is the clinical scenario that perplexes urologists most.

The biggest limitation of all of the studies is the fact that not all patients go on to nephrectomy. In six of the papers we reviewed, zero percent of patients with benign biopsy results went on to nephrectomy, and in others, the numbers were in the single digits. Without a true gold standard to compare benign biopsy results to, the sensitivity and specificity calculations based on these findings are not dependable. Only one study used a time cut-off of two years, after which a benign biopsy result would be considered correct if the mass had not increased in size or become symptomatic. Differing follow-up protocols at different institutions make this determination even more complex. Importantly, some studies have shown malignant masses for which no growth was observed during surveillance. Under this scenario, sensitivity and accuracy values would be inflated because benign biopsies are being considered correct due to the mass’s indolent growth, when in fact they are malignant and the biopsy findings are wrong. In addition to follow-up not being standardized, many authors admit that studies are potentially not long enough to detect whether benign lesions are truly benign; a similar stance has been
raised regarding tumor seeding, noting that studies may not have followed patients long enough to detect seeding if it occurred later on.

The behavior of supposedly benign masses under surveillance is an important aspect of the conclusions of the studies. Because of a lack of nephrectomy follow-up in benign biopsy, authors were forced to use growth-free or symptom-free surveillance periods as the gold standard for an accurate benign biopsy. This is problematic if, as some studies have suggested, growth-free or symptom-free masses can still go on to be malignant. In a 2004 study by Wehle et. al, out of 29 patients enrolled in a surveillance program, the authors reported three cases that went on to nephrectomy because of patient wishes (i.e. not symptoms or substantial growth) and were found to be RCC. For two of these, they reported no growth in 12 months and 38 months of follow-up, respectively. For the third case, they reported a 0.2 cm/year growth rate during 43 months of follow-up. Though this study did have a relatively small sample size, and the majority of patients continued surveillance with no deaths or reports of metastases, the study casts doubt on the use of growth-free surveillance as a true marker of benignity. However, growth-free malignant masses may not be clinically significant if they remain growth-free or grow at a very slow pace. For patients with a limited life expectancy, a mass growing at 0.2 cm/year could be inconsequential even if it is histologically malignant. A longer term study could provide information about the growth of such masses; unfortunately, because of the apprehension of urologists to observe renal masses, there are few surveillance studies in the literature.

All of the case series in our review are published from major academic medical centers throughout the United States and Europe. This adversely affects the generalizability of the study as patients referred to academic medical centers are not an accurate representation of the general population. Similarly, surgeons performing at major academic medical centers often have had
more training and are exposed to higher case volumes than those practicing in community settings. For these reasons, reported outcomes and complications may not accurately portray results in other settings.

**Future Studies**

Here, we propose the design of future studies that would address many of the logistic and epidemiologic shortfalls of the current research. All of the studies in our review were case series reported from major academic medical centers, where patients are more rigorously selected and can introduce unpredictable bias into conclusions. To strengthen generalizability, patients in future studies should be enrolled from population-based registries such as the VA Hospital system or Kaiser because these are more representative of populations across the country. We propose that patients who pose a therapeutic dilemma be pre-specified in some way. For example, those with tumors under 4 cm, age over 65 years old, or those with a pre-defined set of comorbidities that make surgery a less than ideal option. From this point, these patients could be randomized to renal mass biopsy or watchful waiting. In the biopsy group, further treatment would be dictated by biopsy results; type of surgery, biopsy technique, and immunohistochemistry should be standardized to ensure comparability. A standardized follow-up protocol should also be specified; for example, in the first two years, patients would receive CT scans every six months, and if no change is observed, they would subsequently be scanned once a year. Mortality, progression to cancer, and rate of complications could be compared at one, five, and ten year periods. This would be a distinct improvement over the current literature because of the addition of a prospective design, randomization, and a comparison group. From these same population-based registries, studies can report mortality, oncologic outcomes, and complication rate for those patients who do undergo nephrectomy. These findings would be more
generalizable than case series reporting surgical outcomes of highly selected patients at academic medical centers.

**Future Direction**

Renal epithelial cancers are a diverse group with differing natural histories and prognoses due to their own cytogenetic and molecular aberrations\(^\text{65}\). A series of studies have reported improvements in biopsy results with the addition of cytogenetic studies. In a paper by Barocas\(^\text{65}\) et al., the authors combined histopathology with a molecular diagnostic algorithm presented in a previous study\(^\text{66}\) to improve diagnostic accuracy of core biopsy. This algorithm is based on high CA9 expression in clear cell carcinoma, AMACR in papillary RCC, and CLCNKB in chromophobe RCC and oncocytoma\(^\text{66}\). The study reported the diagnostic accuracy of biopsy improved from 83.3% to 95% (57 of 60 biopsies) when histology was combined with the molecular diagnostic algorithm compared to histology alone. The combination increased sensitivity from 87.1% to 100% and improved the negative predictive value from 87.5% to 100%\(^\text{65}\). However, this study has limited generalizability because of the *ex-vivo* sampling of kidney masses. Direct visualization and palpation of masses in the study make extrapolation to a true biopsy scenario difficult. Additionally, the authors report using a 14 gauge needle for biopsy which is substantially larger than the more routinely used 18 gauge needle found in the studies we reviewed. The authors note that there is currently no molecular marker than can distinguish between chromophobe renal cell carcinoma from oncocytoma, which is an often a source of incorrect diagnoses.

A similar study was done in 2006 that combined fluorescent in-situ hybridization (FISH) with histology to improve biopsy accuracy. The authors reported the use of six probes based on the most common mutations in renal cell carcinoma: loss of von Hippel-Lindau (VHL) locus of
chromosome 3p in clear cell RCC, trisomy 7 and 17 in papillary RCC, and losses of parts of chromosomes 10, 13, 17 or 21 in chromophobe RCC. The paper showed that adding FISH led to the correct identification of four additional tumors out of 36 total masses, improving the accuracy of biopsy from 75% to 86%. The study shares weaknesses with the aforementioned study in its ex-vivo design and use of 14 gauge needles. Additionally, FISH is accurate for detecting losses and translocations, but can miss mutations, which can be relevant in clear cell carcinoma. If costs can kept under control, molecular studies could be the future of renal mass management as they will increase diagnostic capabilities and tailor treatment to specific individuals.

Despite increased interest in the topic of RCC, the available literature is composed of low quality studies that lack prospective design, randomization, and generalizability. Keeping this in mind, our review of the literature showed that sensitivity and specificity of renal mass biopsy for differentiating benign and malignant pathology is improving compared to previously published studies and textbooks. We worry, however, that the design of these studies may inflate sensitivity values by leaving failed or indeterminate biopsies out of the denominator. Our study also found a low level of complications and zero reports of needle track seeding, although questions remain about the adequacy of follow-up to assess these findings. The findings of this review show that there is much work left to be done in the area of RCC research. A large, population-based, randomized trial is needed to assess the effect of biopsy on mortality and cancer progression which can then be applied to the general population.
REFERENCES


